
An extension of the Ehrenfeucht-Fraïsse game for first order logics augmented with Lindström quantifiers

Simi Haber
Saharon Shelah

Abstract

We propose an extension of the Ehrenfeucht-Fraïsse game able to deal with logics augmented with Lindström quantifiers. We describe three different games with varying balance between simplicity and ease of use.

Dedicated to Yuri Gurevich on the occasion of his 75th birthday

1 Introduction

The Ehrenfeucht-Fraïsse game [3–6] is an important tool in contemporary model theory, allowing to determine whether two structures are elementary equivalent up to some quantifier depth. It is one of the few model theoretic machineries that survive the transition from general model theory to the finite realm.

There are quite a few known extensions of the Ehrenfeucht-Fraïsse game and in the following we mention a few (this is not a comprehensive list). In [8] Immerman describes how to adapt the Ehrenfeucht-Fraïsse game in order to deal with finite variable logic. Infinitary logic has a precise characterization by a similar game [1, 7]. An extension for fixpoint logic and stratified fixpoint logic was provided by Bosse [2].

Lindström quantifiers were first introduced and studied by Lindström in the sixties [9–12] and may be seen as precursors to his theorem.

The aim of this paper is to present several related extensions of the Ehrenfeucht-Fraïssé game adapted to logics augmented with Lindström quantifiers.

2 The Game

Notation 2.1. (1) Let τ denote a vocabulary. We assume τ has no function symbols, but that is purely for the sake of clearer presentation. τ may have constant symbols.

(2) First order logic will be denoted by \mathcal{L}_{FO} . Along this paper we will look on extensions of first order logic, therefore the logic under discussion will change according to our needs. We shall denote the logic currently under discussion by \mathcal{L} , and we will explicitly redefine \mathcal{L} whenever needed.

(3) Given a vocabulary τ , we use $\mathcal{L}(\tau)$ to denote the *language* with logic \mathcal{L} and vocabulary τ . We will use this notation only when clarity demands, so in fact we may abuse notation and use \mathcal{L} also for the language under discussion.

(4) For even further transparency, all the examples in this manuscript (in particular, all cases of pairs of models to be proved equivalent) will be dealing with simple¹ graphs. Hence (only in examples) we further assume that τ is the vocabulary of graphs denoted henceforth by τ_{GRA} . Explicitly, $\tau_{\text{GRA}} = \{\sim\}$ where \sim is a binary, anti-reflexive and symmetric relation. For the Lindström quantifiers given in examples, we may use vocabularies other than τ_{GRA} .

(5) Let $\mathbf{A}_1, \mathbf{A}_2, \dots$ be classes of models², each closed under isomorphism. The models in \mathbf{A}_i are all τ_i -structures in some relational vocabulary $\tau_i = \{P_{i,1}^{a_{i,1}}, \dots, P_{i,t_i}^{a_{i,t_i}}\}$.

(6) For simplicity, we will assume that each τ_i has an additional relation, $P_{i,0}^1$. This will serve for the formula defining the universe of the model. Formally, all our models will have their domain the entire universe, and the first relation will be a subset defining the domain de facto.

¹An undirected graph with no loops and no double edges is called a *simple graph*.

²I don't see a problem with having infinitely many of these.

(7) We set $a_{i,0} = 1$ for every i .

(8) Each \mathbf{A}_i corresponds to a Lindström quantifier Q_i binding $a_i = \sum_{j=0}^{t_i} a_{i,j}$ variables.

Example 2.2. (1) \mathbf{A}_1 may be the class of commutative groups, in which case τ_1 is consisted of a constant symbol and a ternary relation encoding the group operation.

(2) Another example may be finite Hamiltonian graphs, in which case the vocabulary is the vocabulary of graphs and the class \mathbf{A} will be the set of all finite Hamiltonian graphs (over, say, $[n] = \{1, \dots, n\}$ for any $n \in \mathbb{N}$).

Notation 2.3. Given a vector \bar{x} , we denote its length by $\text{len}(\bar{x})$.

Definition 2.4. We define the quantifier Q_i corresponding to \mathbf{A}_i as follows: Let G be a τ -structure with domain V . For any index i and formulae $\varphi_0(x_0, \bar{y}), \varphi_1(\bar{x}_1, \bar{y}), \dots, \varphi_{t_i}(\bar{x}_{t_i}, \bar{y})$ such that $\text{len}(\bar{x}_j) = a_{i,j}$, the satisfiability of $Q_i x_0, \bar{x}_1, \dots, \bar{x}_{t_i} (\varphi_0(x_0, \bar{b}), \varphi_1(\bar{x}_1, \bar{b}), \dots, \varphi_{t_i}(\bar{x}_{t_i}, \bar{b}))$ is given by

$$\begin{aligned} G \models Q_i x_0, \bar{x}_1, \dots, \bar{x}_{t_i} (\varphi_0(x_0, \bar{b}), \varphi_1(\bar{x}_1, \bar{b}), \dots, \varphi_{t_i}(\bar{x}_{t_i}, \bar{b})) &\iff \\ &(\{x_0 \in V \mid G \models \varphi_0(x_0, \bar{b})\}, \{\bar{x}_1 \in V^{a_{i,1}} \mid G \models \varphi_1(\bar{x}_1, \bar{b})\}, \dots, \\ &\{\bar{x}_{t_i} \in V^{a_{i,t_i}} \mid G \models \varphi_{t_i}(\bar{x}_{t_i}, \bar{b})\}) \in \mathbf{A}_i, \end{aligned}$$

where \bar{b} are parameters.

Remark 2.5. Definition 2.4 requires φ_0 to have exactly one free variable, x_0 (excluding \bar{y} , saved for parameters). However there is not real reason to avoid sets of vectors of any length from serving as the domain of the model defined in the quantifier. We will not discuss this here, but the generalization of the proposed games to this case are straightforward.

Definition 2.6. (1) Let τ be a vocabulary and $\mathcal{L} = \mathcal{L}(\tau)$ be a language. Given two τ -structures G_1, G_2 (not necessarily with distinct universe sets) and two equal length sequences of elements $\bar{x}_1 \in G_1, \bar{x}_2 \in G_2$, we say that (G_1, \bar{x}_1) and (G_2, \bar{x}_2) are *k-equivalent with respect to \mathcal{L}* if for any formula $\varphi(\bar{x}) \in \mathcal{L}$ of quantifier depth at most k one has

$$G_1 \models \varphi(\bar{x}_1) \iff G_2 \models \varphi(\bar{x}_2).$$

-
- (2) When considering only one model, that is, when we take $G = G_1 = G_2$, we refer to the equivalence classes of this relation in the domain of G simply by *the (a, k, G) -equivalence classes* (or just equivalence classes when the context is clear enough).

Notice that unions of (a, k, G) -equivalence classes are exactly the definable sets of a -tuples of elements in $\text{dom}(G)$ using \mathcal{L} -formulas of quantifier depth at most k .

Example 2.7. Let \mathcal{L} be the first order language of graphs, $\mathcal{L} = \mathcal{L}_{\mathcal{FO}}(\tau_{\text{GRA}})$, and let $G = (V, E)$ be a graph. If G is simple then the $(1, 0, G)$ -equivalence classes are V and \emptyset . If $|V| > 1$ then the $(1, 1, G)$ -equivalence classes are³ the set of isolated vertices in G , the set of vertices adjacent to all other vertices and the set of vertices having at least one neighbor and one non-neighbor (some of which may be empty of course).

Notation 2.8. We denote the logic obtained by augmenting the first order logic with the quantifiers Q_1, Q_2, \dots by $\mathcal{L} = \mathcal{L}[Q_1, Q_2, \dots]$.

Example 2.9. Consider $\mathcal{L} = \mathcal{L}[Q_{\text{HAM}}](\tau_{\text{GRA}})$, where Q_{HAM} stands for the “Hamiltonicity quantifier” (corresponding to the class of graphs containing a Hamiltonian cycle — a cycle visiting each vertex precisely once). Let G be a graph. Then the set of all vertices x for which all of the graphs⁴ $G[N_G(x)], \overline{G}[N_G(x)], G[N_{\overline{G}}(x)], \overline{G}[N_{\overline{G}}(x)]$ are Hamiltonian is an example of a $(1, 1, G)$ -equivalence class with respect to $\mathcal{L}[Q_{\text{HAM}}]$. The set of vertices with degree exactly two is a union of $(1, 1, G)$ -equivalence classes, as can be seen by⁵

$$\varphi(x) = Q_{\text{HAM}}x_0, x_1, x_2 (x_0 \sim x, x_1 \neq x_2).$$

2.1 Description of the first game

Before describing the game, we need the following definition:

Definition 2.10. Let τ be a vocabulary, \mathcal{L} a language over that vocabulary (not necessarily first order) and G a model of τ . Additionally, let $M =$

³The atomic sentences appearing in $\varphi(x)$ are $x = y$ and $x \sim y$.

⁴Here $N(x) = \{y \in V \mid x \sim y\}$ is the neighborhood of x in G , $G[U]$ where $U \subseteq V$ is the graph induced on U and \overline{G} is the complement of G .

⁵ φ expresses: “the complete graph $K_{d(x)}$ is Hamiltonian” which is true when $d(x) > 2$ and false when $d(x) = 2$ (we may treat K_0 and K_1 separately, if needed).

(S', R'_1, \dots, R'_t) be a model of another vocabulary τ' . A *copy of M in G* is a tuple (S, R_1, \dots, R_t) such that

- (1) S is a subset of $\text{dom}(G)$ with the same cardinality as $\text{dom}(M) = S'$ (where $\text{dom}(G)$ is the universe or underlying set of G).
- (2) R_1, \dots, R_t are relations over S , such that each R_j has the same arity as R'_j .
- (3) (S, R_1, \dots, R_t) is isomorphic to (S', R'_1, \dots, R'_t) .

If in addition the following holds

- (4) S is union of $(1, k, G)$ -equivalence classes, and each relation R_j of arity a_j is a union of (a_j, k, G) -equivalence classes;

we say that a copy of M in G is *k -induced by \mathcal{L}* . When k and \mathcal{L} can be clearly determined by the context, we may omit mentioning one of them, or both.

We may now define the first game.

Definition 2.11. Let G_1 and G_2 be two models with domains V_1 and V_2 respectively. Let $k \geq 0$ an integer and $\bar{c}_\ell = (c_\ell^1, \dots, c_\ell^r) \in V_\ell^r$ two finite sequences. We define the game⁶ $\text{EFL}_1[G_1, G_2, \bar{c}_1, \bar{c}_2; k]$. There are two players, named ISO and AIS. The game board is the models G_1 and G_2 plus the sequences \bar{c}_ℓ and there are k rounds. Each round is divided into two parts, and each part consists of two sub-rounds. The game is defined recursively. If $k = 0$, then if the mapping $c_1^i \rightarrow c_2^i$ is an isomorphism, then ISO wins, otherwise AIS wins.

When $k > 0$ then first AIS plays. He picks one of the models G_1 or G_2 (denoted henceforth by G_ℓ) and a quantifier Q_i (or the existential quantifier⁷). Next AIS picks a model $M \in \mathbf{A}_i$, and embeds it into G_ℓ in a manner that preserve $(k-1, G_\ell)$ -equivalence classes. That is, AIS picks a tuple $(S_\ell, R_{\ell,1}, \dots, R_{\ell,t_i})$ that is a copy of M in G which is $(k-1)$ -induced by \mathcal{L} enriched with r constants having values \bar{x}_ℓ . If AIS can not find such an embedding, he loses⁸. Implicitly AIS claims that ISO can not find a matching induced copy of a model from \mathbf{A}_i .

⁶We will describe a few variants, hence the subscript.

⁷In this case, $\mathbf{A}_\exists = P(V) \setminus \{\emptyset\}$, so AIS may choose any non-empty subset S_ℓ of V_ℓ .

⁸We will consider only logics stronger than first-order, hence the existential quantifier is always assumed to be at AIS' disposal and he will never lose in this manner.

Second, ISO responds by choosing a model M' from \mathbf{A}_i (M' may not necessarily be the same as M), and then picking an induced copy of M' in $G_{3-\ell}$ which we naturally denote by $(S_{3-\ell}, R_{3-\ell,1}, \dots, R_{3-\ell,t_i})$. She is implicitly claiming that her choices match the picks of AIS, that is, each $R_{3-\ell,j}$ (or $S_{3-\ell,j}$) is a union of $(a_{i,j}, k-1, G_{3-\ell})$ -equivalence classes defined by the same formulas as the formulas defining the $(a_{i,j}, k-1, G_\ell)$ -equivalence classes of which $R_{\ell,j}$ is made. If ISO can not complete this part she loses. This ends the first part of the round.

In the second part of the round AIS chooses $m \in \{1, 2\}$ and $0 \leq j \leq t_i$. He then picks $(c_m^{r+1}, \dots, c_m^{r+a_{i,j}}) \in R_{m,j}$ (implicitly challenging ISO to do the same). Finally ISO picks $(c_{3-m}^{r+1}, \dots, c_{3-m}^{r+a_{i,j}}) \in R_{3-m,j}$ and they move on to play

$$\text{EFL}_1[G_1, G_2, (c_1^1, \dots, c_1^r, c_1^{r+1}, \dots, c_1^{r+a_{i,j}}), \\ (c_2^1, \dots, c_2^r, c_2^{r+1}, \dots, c_2^{r+a_{i,j}}); k-1].$$

This ends the second part and the round. Since k goes down every round, the game ends when $k = 0$, as described above.

Given the description above, the following should be self-evident:

Lemma 2.12. *Let $\mathcal{L} = \mathcal{L}[Q_1, Q_2, \dots](\tau)$ be a language over some vocabulary τ where Q_1, Q_2, \dots are Lindström quantifiers and let G_1, G_2 be two τ -structures. Then, ISO has a winning strategy for $\text{EFL}_1[G_1, G_2, \emptyset, \emptyset; k]$ if and only if for any sentence $\varphi \in \mathcal{L}$ of quantifier depth at most k*

$$G_1 \models \varphi \iff G_2 \models \varphi.$$

2.2 A game where definability is not forced

While the claim of Lemma 2.12 is satisfying, it may be hard to put into use since it takes finding unions of $(a, k-1, G)$ -equivalence classes for granted, being a rule of the game. This might hinder strategy development and we would like to describe another game with looser rules, denoted EFL_2 .

In this version the players are not bound to choosing unions of $(a, k-1, G)$ -equivalence classes when picking a copy of the chosen model (hence we call their action “picking a copy of M in G_ℓ ”, omitting the “induced” part). That is, we omit requirement 4 in Definition 2.10. It falls to the other player to

check that indeed every relation is a union of the relevant equivalence classes. A general round now goes as follows:

AIS picks a graph $G_\ell \in \{G_1, G_2\}$ and a quantifier Q_i (or, as before, the existential quantifier). Next AIS picks a model $M \in \mathbf{A}_i$ and picks a copy of M in G_ℓ . His implicit claim now includes the claim that each of the relations he chose is a union of $(a_{i,j}, k-1, G_\ell)$ -equivalence classes with respect to \mathcal{L} enriched with r constants having values \bar{c}_ℓ .

ISO can respond in two different ways — she can “accept the challenge” (as she did in EFL_1), or attack the second part of the claim of AIS. That is, she can do one of the following:

- (1) Accept. In this case she chooses $M' \in \mathbf{A}_i$ and picks a copy of M' in $G_{3-\ell}$. Implicitly she is claiming that her choices matches the choices of AIS. That is, the set of vertices $S_{3-\ell}$ and each of the relations defined on it are a union of the $(a_{i,j}, k-1, G_{3-\ell})$ -equivalence classes corresponding⁹ to the ones that AIS picked. This ends the first part of the round.

AIS may continue in a two different ways.

- (a) Reject the fact that $S_{3-\ell}$ or one of the relations picked by ISO is a union of equivalence classes. In order to settle this, we recursively use EFL_2 :

Again, we let $a = a_{i,j}$ be the arity of the allegedly invalid relation $R_{3-\ell,j}$. AIS picks two a -tuples, $(c^{r+1}, \dots, c^{r+a}) \in R_{3-\ell,j}$ and $(c'^{r+1}, \dots, c'^{r+a}) \in V_{3-\ell}^a \setminus R_{3-\ell,j}$, and they move on to play

$$\text{EFL}_2[G_{3-\ell}, G_{3-\ell}, (c_{3-\ell}^1, \dots, c_{3-\ell}^r, c^{r+1}, \dots, c^{r+a}), \\ (c_{3-\ell}^1, \dots, c_{3-\ell}^r, c'^{r+1}, \dots, c'^{r+a}); k-1].$$

with exchanged roles (since this time AIS claims the two tuples are actually $(a, k-1, G_{3-\ell})$ -equivalent).

- (b) Reject the fact that ISO’s choice matches his choice (as he did in EFL_1). In this case he picks a relation $P_j \in \tau_i$ and an $a_{i,j}$ -tuple

⁹We say that E_1 , an (a, k, G_1) -equivalence class of a -tuples in G_1 corresponds to E_2 — a set of a -tuples in G_2 if for any $\bar{x}_1 \in E_1$ and $\bar{x}_2 \in E_2$ one has

$$G_1 \models \varphi(\bar{x}_1) \Leftrightarrow G_2 \models \varphi(\bar{x}_2)$$

for any $\varphi \in \mathcal{L}$ of quantifier depth at most k .

of elements from S_ℓ (or one element if he challenges her choice of $S_{3-\ell}$). Denote his choice by $(c_\ell^{r+1}, \dots, c_\ell^{r+a_{i,j}}) \in S_\ell$. ISO responds by picking another a -tuple $(c_{3-\ell}^{r+1}, \dots, c_{3-\ell}^{r+a_{i,j}}) \in S_{3-\ell}$, and they move on to play

$$\text{EFL}_2[G_1, G_2, (c_1^1, \dots, c_1^r, c_1^{r+1}, \dots, c_1^{r+a_{i,j}}), \\ (c_2^1, \dots, c_2^r, c_2^{r+1}, \dots, c_2^{r+a_{i,j}}); k-1].$$

- (2) Reject. In this case ISO wants to prove that S_ℓ or one of the relations picked by AIS is not a union of equivalence classes. We continue similarly to case 1.(b):

Let $a = a_{i,j}$ be the arity of the allegedly relation, $R_{\ell,j}$, splitting an equivalence class. ISO picks two a -tuples, $(c^{r+1}, \dots, c^{r+a}) \in R_{\ell,j}$ and $(c'^{r+1}, \dots, c'^{r+a}) \in V_\ell^a \setminus R_{\ell,j}$, and they move to play

$$\text{EFL}_2[G_\ell, G_\ell, (c_\ell^1, \dots, c_\ell^r, c_\ell^{r+1}, \dots, c_\ell^{r+a}), \\ (c_\ell^1, \dots, c_\ell^r, c_\ell'^{r+1}, \dots, c_\ell'^{r+a}); k-1].$$

this time keeping their original roles.

For any two models G_1 and G_2 , constants \bar{c}_1, \bar{c}_2 and $k \in \mathbb{N}$, whoever has a winning strategy for $\text{EFL}_1[G_1, G_2, \bar{c}_1, \bar{c}_2; k]$ has a winning strategy for $\text{EFL}_2[G_1, G_2, \bar{c}_1, \bar{c}_2; k]$. Hence the parallel of Lemma 2.12 is true for EFL_2 as well.

While we got the benefit of in-game validation of the equivalence classes integrity claims, EFL_2 is not easy to analyze in applications because the game-board and players role change over time. We amend this in the last suggested version of the game.

2.3 A game with fixed game-board and roles

The last version, denoted EFL_3 , forks from EFL_2 in two places.

Definition 2.13. We define EFL_3 like EFL_2 except that:

- (1) First, assume the game reaches step 2., where ISO wants to prove that AIS's chose a relation $R_{\ell,j}$ splitting an equivalence relation. In this case the first part of the round ends immediately and the second part goes as follows:

ISO chooses two $a_{i,j}$ -tuples, $\bar{c}_{\ell,1}$ from $R_{\ell,j}$ and $\bar{c}_{\ell,2}$ from the complement of $R_{\ell,j}$. She then pick another $a_{i,j}$ -tuples from $G_{3-\ell}$, denoted $\bar{c}_{3-\ell}$. Spoiler then picks one of $\bar{c}_{\ell,1}$ or $\bar{c}_{\ell,2}$ and they move on to play EFL_3 with $\bar{c}_{3-\ell}$ concatenated to the constants of $G_{3-\ell}$ and AIS's choice concatenated to the constants of G_ℓ , and $k-1$ moves. They keep their roles and the game-board is still G_1 and G_2 .

If ISO can not find a matching tuple in $G_{3-\ell}$, she can not disprove the integrity claim of AIS, but it does not matter as G_1 and G_2 are not k -equivalent and she is bound to lose anyway.

Notice that in this case the first part of the round had only AIS playing, and in the second part ISO played first.

- (2) The second (and last) change from EFL_2 happens when the game is in step 1a. In this case AIS wants to prove that ISO's choice of at least one relation $R_{3-\ell,j}$ is splitting an equivalence relation. In this case AIS picks a tuple $\bar{c}_{3-\ell}$ (from the suspicious equivalence class) in $G_{3-\ell}$ that is not in $R_{3-\ell,j}$ and challenges ISO to find a matching tuple \bar{c}_ℓ in $G_{3-\ell}$ that is not in $R_{\ell,j}$. They move on to play EFL_3 with these choices and $k-1$ moves. Again both roles and game-board remain as was. Notice that the game flow in this case is actually the same as the game flow in 1b.

As before, it is easy to convince oneself that the claim of Lemma 2.12 is still valid. We repeat it here:

Lemma 2.14. *ISO has a winning strategy for $\text{EFL}_3[G_1, F_2, \emptyset, \emptyset; k]$ if and only if for any sentence $\varphi \in \mathcal{L}$ of quantifier depth at most k*

$$G_1 \models \varphi \iff G_2 \models \varphi.$$

3 Summary

We have presented three equivalent variants of the celebrated Ehrenfeucht-Fraïssé game adapted to deal with logics extended by Lindström quantifiers. We believe EFL_3 may be easier to analyse than direct quantifier elimination and it is our hope that it will find applications.

References

- [1] J. Barwise. On moschovakis closure ordinals. *The Journal of Symbolic Logic*, 42(2):292–296, June 1977.
- [2] U. Bosse. An “Ehrenfeucht-Fraïssé game” for fixpoint logic and stratified fixpoint logic. In E. Börger, G. Jäger, H. K. Büning, S. Martini, and M. M. Richter, editors, *Selected Papers from the 6th Workshop on Computer Science Logic*, CSL ’92, pages 100–114, London, UK, Sept. 1993. Springer-Verlag.
- [3] A. Ehrenfeucht. An application of games to the completeness problem for formalized theories. *Fundamenta Mathematicae*, 49:129–141, 1961.
- [4] R. Fraïssé. Sur une nouvelle classification des systèmes de relations. *Comptes Rendus*, 230:1022–1024, 1950.
- [5] R. Fraïssé. *Sur quelques classifications des systèmes de relations*. PhD thesis, University of Paris, 1953. Published in [6].
- [6] R. Fraïssé. Sur quelques classifications des systèmes de relations. *Publications Scientifiques de l’Université d’Alger, series A*, 1:35–182, 1954.
- [7] N. Immerman. Upper and lower bounds for first order expressibility. *Journal of Computer and System Sciences*, 25(1):76–98, June 1982.
- [8] N. Immerman. *Descriptive Complexity*. Graduate Texts in Computer Science. Springer, New York, 1998.
- [9] P. Lindström. First order predicate logic with generalized quantifiers. *Theoria*, 32(3):186–195, Dec. 1966.
- [10] P. Lindström. On characterizability in $L_{\omega_1\omega_0}$. *Theoria*, 32(3):165–171, Dec. 1966.
- [11] P. Lindström. On relations between structures. *Theoria*, 32(3):172–185, Dec. 1966.
- [12] P. Lindström. On extensions of elementary logic. *Theoria*, 35(1):1–11, Apr. 1969.